Archipelago Appliance Evolution

1. Introduction

Archipelago seeks to improve the performance and security of computer networks to make it easier and less costly for scientists to move, store, and analyze their data. Archipelago accomplishes these goals through the use of small quantities of Software Defined Networking (SDN) nodes deployed within traditional campus networks to form a hybrid architecture. A portion of the Archipelago architecture leverages low-cost SDN appliances developed utilizing common off-the-shelf components. This document describes the evolution of these appliances with each generation providing added performance and resiliency while minimizing cost. Before reviewing the evolution of the appliance over time, we present overall design goals within §2.

2. Design

While Archipelago can utilize purpose-built next-generation SDN switching hardware, we also wanted to provide a low-cost yet high-performance appliance in support of improving the adoption of SDN within the campus architecture. To develop a low-cost high-performance SDN appliance, we wanted to leverage commodity off-the-shelf components with open-source software. At the same time, we wanted to develop an appliance that had a physical form factor similar to that of existing network forwarding elements, such as switches, to make it easy to deploy alongside traditional infrastructure. Moreover, we wanted to ensure that the forwarding capabilities of this appliance would support the transfer of large volumes of data in support of research. Our design characteristics are summarized as follows:

- Short-depth 1RU chassis (19” wide, 1.75” tall, < 20” deep)
- Front I/O - network ports on the front of the chassis, like a traditional switch
- Rear power – resilient and hot-swap if possible
- Low noise, robust and resilient cooling – front-to-back airflow (hot-swap desirable)
- Smart Network Interface Card (SmartNIC) in support of forwarding acceleration

The remaining sections within this document describe the evolution of the Archipelago appliance as it evolved over time.

3. First Generation Appliance

Locating a chassis with the above characteristics proved to be challenging. The first suitable chassis we could identify was the Supermicro SuperChassis 515-505 as shown in Figure 3.1 below:

Figure 3.1 - First Generation Appliance Chassis – Supermciro SuperChassis 515-505

Image Courtesy of Supermicro

The chassis features two front-mounted PCIe 3.0 x16 slots, temperature-controlled fans for low noise, and a single rear-mounted power supply. The 17" depth made this an ideal chassis for mounting in 2- or 4-post racks. This appliance was outfitted as follows:

- Supermicro SuperChassis 515-505
- 4 temperature controlled fans
- Supermicro X11SPW-TF Motherboard
- Intel Xeon Silver 4110 (8-Core 2.1GHz, 11M Cache) - 16 CPU Threads
- 2x 32-GB DDR4 ECC Registered Memory
- 1 x 512GB - Samsung 970 PRO -- M.2 NVMe SSD
- 8-pin to 2 PCI-e 8 (6+2) pin PCI-e Power Connector for GPU
- 1x Nvidia single-slot Quadro RTX 4000 GPU (8GB GDDR6)
- 1x Netronome Agilio 2x25Gb/s or 2x40Gb/s SmartNIC

This chassis worked well in our lab tests although, as mentioned in §2, there were still some items missing such as resilient hot-swap power supplies.

4. Second Generation Appliance

We conducted power analysis with the GPU, NPU and CPU running at 100% load within the chassis highlighted within §3 and found the new Supermicro SuperChassis 515-R407 featuring resilient hot-swap power supplies to be a candidate chassis for an Archipelago appliance and it became our second-generation chassis as shown in Figure 4.1:

Figure 4.1 - Second Generation Appliance Chassis – Supermciro SuperChassis 515-R407

Image Courtesy of Supermicro

The chassis features two front-mounted PCIe 3.0 x16 slots, temperature-controlled fans for low noise, and dual resilient hot-swap rear-mounted power supplies. The 17” depth made this an ideal chassis for mounting in 2- or 4-post racks. This appliance was outfitted as follows (changes from the first-generation chassis are *italicized*):

- *Supermciro SuperChassis 515-R407*
- *2 additional temperature controlled fans (6x total)*
- Supermicro X11SPW-TF Motherboard
- *Intel Xeon® Silver® 4208 (8-Core 2.1GHz, 11M Cache) - 16 CPU Threads*
- 2x 32-GB DDR4 ECC Registered Memory
- 1 x 512GB - Samsung 970 PRO -- M.2 NVMe SSD
- 8-pin to 2 PCI-e 8 (6+2) pin PCI-e Power Connector for GPU
- 1x Nvidia single-slot Quadro RTX 4000 GPU (8GB GDDR6)
- 1x Netronome Agilio 2x25Gb/s or 2x40Gb/s SmartNIC

This chassis worked well in our lab tests including running the GPU, NPU and CPU at 100% load on single power supply although, as mentioned in §2, there were still some items missing such as hot-swap fans. Additionally, the first two appliance generations featured a single non-resilient NVMe SSD drive. With only a single M.2 port on the motherboard, we considered options such as PCIe to M.2 converter card, but did not want to lose a front-facing PCIe slot. These challenges were addressed in the third-generation appliance presented below.

5. Third Generation Appliance

Supermicro introduced a newer SuperServer SYS-110P-FRN2T that we selected for our third-generation appliance. As a system, the motherboard and other components came pre-integrated into the chassis at the factory minimizing the amount of customization and special component stock required by Supermicro integrators. Moreover, the chassis used with SuperServer SYS-110P-FRN2T features hot-swap resilient fans which were missing in the first two generations of appliances. The system is shown In Figure 5.1 below:

Figure 5.1 - Third Generation Appliance Chassis – Supermciro SuperServer SYS-110P-FRN2T

Image Courtesy of Supermicro

Figure 5.2 - Third Generation Appliance Chassis – Rear View

Image Courtesy of Supermicro

The system features two front-mounted *PCIe 4.0* x16 slots, hot-swap temperature-controlled fans for low noise, and dual resilient hot-swap rear-mounted power supplies. The *16" depth* made this an ideal chassis for mounting in 2- or 4-post racks. We have developed two standard configurations for this appliance described within the subsections below. This system addresses the shortcomings of the original chassis designs and provides additional resiliency and performance when compared to the prior appliance generations.

5.1 SDN Controller / Data Plane Persona

This persona is used for the standard SDN controller and data plane use cases for Archipelago as follows (changes from the second-generation chassis are *italicized*):

- *Supermciro SuperServer SYS-110P-FRN2T*
- *3 additional temperature controlled fans (5x total)*
- *Supermicro Super X12SPW-TF Motherboard*
- *Intel Xeon Silver 4309Y Processor 8-Core 2.8GHz 12MB Cache (105W)*
- *4 x 16GB PC4-25600 3200MHz DDR4 ECC RDIMM*
- *2 x 480GB Micron 5300 MAX Series 2.5" SATA 6.0Gb/s Solid State Drive*
- 1 x 512GB - Samsung 970 PRO -- M.2 NVMe SSD
- *Supermicro CBL-PWEX-1040 - 8pin EPS CPU to 6/6+2pin GPU Power Adapter*

- Optional: 1x NVIDIA RTX A4000 - 16GB GDDR6 - PCIe 4.0 x16 - Active Cooling (4xDP)
- Optional: 1x Netronome Agilio 2x25Gb/s or 2x40Gb/s SmartNIC

We have selected this system for use with the pilot Archipelago deployment at Duke as described in our related document "archipelago-final-use-cases-v1.1". We have also elected to use this architecture for several use cases at Duke BEYOND Archipelago including the following:

- Stratum-2 NTP appliances with Intel X550 NIC supporting hardware timestamping
- PerfSONAR deployments as a part of the NCShare Science DMZ (OAC-2201525)

5.2 Sensor / Cybersecurity Persona

This persona is used for projects outside of Archipelago such as for sensor nodes within our SDN driven monitoring fabric as discussed within the extension section of "archipelago-final-use-cases-v1.1". In this deployment scenario, traffic from strategic vantage points is collected with passive optical taps or port mirroring sessions (SPAN ports) on switches/routers. The traffic is sliced according to policy by purpose-built SDN switches and delivered to sensor nodes for flow log generation (ex: IPFIX, NetFlow) and IDS (Intrusion Detection System) analysis. The traffic is presented to these nodes and a Mellanox ConnectX-5 100Gb/s NIC is then virtualized with PCIe SR-IOV to present the traffic to containers of interest. Given that software is processing large volumes of data in this use case, we opted for a slightly different configuration that features 2x the number of CPU cores and 2x the RAM as follows (changes from the appliance presented in §5.1 are *italicized*):

- Supermciro SuperServer SYS-110P-FRN2T
- 3 additional temperature controlled fans (5x total)
- Supermicro Super X12SPW-TF Motherboard
- *Intel Xeon Gold 6326 Processor 16-Core 2.9GHz 24MB Cache (185W)*
- *8 x 16GB PC4-25600 3200MHz DDR4 ECC RDIMM*
- 2 x 480GB Micron 5300 MAX Series 2.5" SATA 6.0Gb/s Solid State Drive
- 1 x 512GB - Samsung 970 PRO -- M.2 NVMe SSD
- Supermicro CBL-PWEX-1040 - 8pin EPS CPU to 6/6+2pin GPU Power Adapter
- *Mellanox 100-Gigabit Ethernet Adapter ConnectX-5 Ex EN MCX516A (2x QSFP28) - PCIe 4.0 x16*

This appliance is deployed in a prototype version of the SDN monitoring fabric as part of MISTRAL - Massive Internal System Traffic Research Analysis and Logging (OAC-22319). Moreover, this appliance has been deployed as a part of the NCShare Science DMZ (OAC-2201525) for IDS and flow monitoring.